Theory-Based User Modeling for Personalized interactive Information Retrieval

Asad Ullah
Institute for Research in Applicable Computing
University of Bedfordshire
, LU1 3JU, UK
Asadullah.asadullah1@study.beds.ac.uk

Haiming Liu
Department of Computer Science and Technology
University of Bedfordshire
LU1 3JU, UK
Haiming.liu@beds.ac.uk

ABSTRACT
In an effort to improve users' search experiences during their information seeking process, providing a personalized information retrieval system is proposed to be one of the effective approaches. To personalize the search systems requires a good understanding of the users. User modeling has been approved to be a good method for learning and representing users. Therefore many user modeling studies have been carried out and some user models have been developed. The majority of the user modeling studies applies inductive approach, and only small number of studies employs deductive approach. In this paper, an EISE (Extended Information goal, Search strategy and Evaluation threshold) user model is proposed, which uses the deductive approach based on psychology theories and an existing user model. Ten users' interactive search log obtained from the real search engine is applied to validate the proposed user model. The preliminary validation results show that the EISE model can be applied to identify different types of users. The search preferences of the different user types can be applied to inform interactive search system design and development.

CCS Concepts
•Human-centered computing → HCI design and evaluation methods; User models;

•Information systems → Users and interactive retrieval; Personalization;

Keywords
Information retrieval, User modeling, Personalization

1. INTRODUCTION
Search engines e.g. (Google and Bing) are the core online technologies, which are widely used by people for information seeking. Information retrieval is to query the search engine with one or multiple iterations [28]. The returning results can be relevant to one group of users but not to the others due to the users' different information needs [26]. Therefore, to understand the users' needs and preferences and to personalize the search process becomes vital to deliver good search experience to the users. To enable the personalized search, different methods are proposed. For example, making query suggestions to the users [20, 12], Users' profiling [14] and modeling users' interest [17]; and Providing different results to different users based on the user short and long term search behaviors [2]. To understand users' needs and preferences, user modeling is an effective approach. The existing user modeling research mainly applies the inductive approach, which means that the development of the user model is informed by the analysis results of the user's interaction data. However, the results generated by this approach are often not validated by any users and theories. Therefore, a deductive approach is employed in this paper to design and develop an effective theory-based user model for personalized interactive search. The model is preliminarily validated using ten users' real life search logs from a search engine.

This paper is divided into the following sections. Section 2 is related work to the current study, section 3 proposes the model, Section 4 describes methodology and Section 5 reports experimentation set up and results.

2. RELATED WORK
In this section, some related user behavior theories, psychology theories and user models will be reviewed. The proposed model by Wilson in 1981 [31] is one of the first model on information seeking behavior. The model splits into two parts: one is how information needs are initiated; the other is what attainment barriers are. The users' information needs often rely on the surrounding environment, psychological needs and Social cultural. Political and economic needs also affect the information needs and search strategy. The model proposed by Wilson in 1996 [32] is a more refined model on the basis of using numerous skills rather than just information science skills, e.g. decision making, innovation, and consumer research. Another model developed by Ellis [6] explains step by step approach involved in information seeking. The model includes six steps, namely, starting, chaining, browsing, differentiating, extracting and finally ending. Information Search Process (ISP) model proposed by Kuhlthau [13] includes six stages, namely, initiation, selection, exploration, formulation, collection and presentation. Each of the six stages also explain the emotions behind the behind the scene [8]. These models provide useful theoretical framework to personalized information seeking process and valuable contribution to the knowledge. However, whilst these models explain the human perspective of interactive information seeking, there is a lack of validation based on users' interaction data. This study will not only propose the theory-based user model for interactive

information seeking, but also validate the user model using the users' real life search log.

The above user models are developed based on a deductive approach (theory-based). The following section will introduce user models developed based on the inductive approach (user data analysis based). In our view, a combination of the two is need for user model development.

Query Suggestion is one of the popular approaches for personalization [3]. For example context aware query suggestions, [3] and identification of different aspects of queries, [9, 10, 12]. User profiling is another method to find out users' interests at different level [30]. Other studies focus on user long term search history and applying probabilistic model on the search history [25]. In other studies the short and long term interest are combined to optimize the effect of personalization [2]. Similar study by [15] integrates long term and short term history to build a user profile. Groups of researcher have done a study on the user unique information goals. They have found that search engines can satisfy overall intentions of user but cannot address their unique goals. [26]. Another study [27] models users decision points where they explain the flow of user from choosing search engine to summarizing results and selecting result or moving into another search provider if not satisfied. Our study will induce both information goal and decision making in our user model, which will be supported by psychology theories and validated by the real life user search interaction data.

In this paper, a user classification model called ISE (Information Goal, Search strategy and Evaluation threshold) [16] is modified in a new context. A new user model called EISE (Extend ISE) will be proposed and validated based on the established phycology theories and users' search interaction date. The ISE user classification model is developed based on Information Foraging Theory (IFT) [21]. IFT was originally derived from Optimal Foraging Theory (OFT) [29]. OFT found a pattern when animals hunting for prey, how animal chose their food depends on their environment and abundance of food, for example, the decision on whether they should eat or leave the food and move to a different place [23]. Pirolli and card adapted the theory to the information world. They found lots of similarities in animal hunting and human information seeking. Based on this, they proposed information foraging theory (IFT) [21]. IFT consists of information scent model, information patch model, and information diet model [21, 22]. The ISE user classification model groups user into three categories and six characteristics namely information goal (Fixed and Evolving), Search Strategy (Cautious and risky), evaluation threshold (Precise and weak) [21]. The information goal of ISE model is developed based on the information sent model of IFT. According to IFT user with strong clues normally have fixed information goals, otherwise the user will be considered to have evolving information goal that is also called exploratory search [16]. The search strategy is developed based on the information patch model of IFT. Cautious users will move around very carefully to find relevant information and when risky users will move around between patches a lot [15]. Finally evaluation threshold is developed based on information diet model. Decision making process is heavily involved at this stage. Normally precise user will be picky when selecting results, whilst weak user will be satisfied with the result easily [16].

3. PROPOSED MODEL – ESIE

Based on the ISE model [16] introduced in the last section, users can be grouped based on six characteristics and of three search behavior categories (Table 1). Whilst the ISE model is proven to be useful to model users for their search preferences, the limitation of the model is that there is a lack of the operational definitions of each characteristic for more precise user modeling. In this section, the operational definitions of the six characteristics of the ISE model will be enriched on the bases on psychology theories. Our assumption is that the users' normal preferences and behaviors should be similar to the users' preference and behaviors when searching. Psychology theories will be reviewed and employed in this section to extend ISE model (EISE). The following subsections will describe how the Psychology theories are applied to extend the ISE mode.

Table 1: The three categories and six characteristics of ISE user classification model.

Categories	Characteristics	
Information Goal	Fixed	Evolving
Search Strategy	Cautious	Risk
Evaluation Threshold	Precise	Weak

3.1 Information Goal

This section will describe the two characteristics of Information Goal, fixed information goals and evolving information goals and extend these characteristics with two mind-set theory (Fixed and Growth mind-sets) [5]. According to Carol dweck two types of mind sets exists fixed mind set and growth mind set [5]. The two mind-set theory has different usage in psychology and personality building. She prefers the growth mind set over the fixed mind set but this is not our research concern. In this study the adapted characteristics of those mind sets will be engaged to fit in EISE model characteristics. The study will investigate the existence of those behaviors in the information seeking process. Our research dose not focus on cultivating certain type mind set but to provide a personalized search experience to the certain type of mind set. Table 2 and Table 3 will present both mind sets mapped to the Fixed and Evolving Information Goal of the ISE model. The obtained analogy from the two mind theory and the derived operational definitions from the analogy will explicitly help us to distinct between users. The following sections will explain the differences from the two mind sets and the similarity between the two minds sets theory and the Information Goal of the ISE model.

3.1.1 Fixed Information Goal

Detail explanation of information goals are delivered in the above sections. In this section the focal point will be one of the characteristic of information goal. Correlation between fixed mind –set and fixed information goal can help us to build a resilient model established psychology theory. From analogy of the theory we can produce an operational definition to distinguish between user behaviors. Operational definitions are a set of rules to quantify user activities.

In the left column of the Table 2 presents Fixed mind-set theory of the two minds theory. The right column of Table 1 describes the search analogy of the Fixed mind-set theory.

Table 2: Fixed mindset theory and its search analogy.

Fixed mind-set	Analogy
The Fixed mind-set people believe to be smarter and can do anything intelligently. People with Fixed mind-set avoid challenges, criticism, obstacles, and effort and try to escape by the shortest possible way and because of their behavior and their minds are fixed to certain amount of understanding and he will miss enormous amount intellect power in consequence of missing a better solution to the problem [5]	The search analogy of this theory can be linked with the Fixed information goal in the ISE model, which is that users with fixed information goal already has something in mind so a user will perform a very limited interaction with the search system. The user will also avoid obstacles which mean a very superficial search process with no effort to details. Finding the results while browsing and because of this user can miss important information.

From the above conclusion of Table 1 the following operational definitions can be obtained from the description of the theory and the search analogies:

- **Less number of query iterations.**
  Query iterations are alterations to the query by user. In the case of fixed information goals the users will have less number of subset, superset and overlaps.
- **Use small number of fixed jumps.**
  Fixed jumps are the types of query jumps used when there are no changes to the information goals user. In this case user will have less number of fixed jumps because the user is not interested to look deeper in the search engine.
- **Use small number of history.**
  History is an example of queries that are used in sessions and between the sessions. User with fixed information goals will be satisfied with first attempts of information seeking. The user will not return back refine his search.

The detail functions of these operational definitions are explained in the experimental set up in section 4.

3.1.2 Evolving Information Goal

The characteristics of growth mind set can be mapped with evolving information goals. According to two mind set theory people with growth mind can exceed more than fixed mind set but in the model will only adapt the characteristics of the growth mindset to endorse our model. The preference of one mind set has no significance on other mind. Because the study is eager to identify the variance between fixed and evolving information goals.

Table 3 shows the summary of the growth mindset theory in the left column, and the right column shows the analogy of the Growth mindset theory in search scenario.

Table 3: Growth mind-set and its search analogy.

Growth mind-set	Analogy
The users with Growth mind-set are keen to learn. They welcome accept challenges and criticisms. They looked forward to see obstacles and efforts as productive opportunities. User with Growth mind-set can learn more about the topic and knowledge by exploring and digging the problem. [5]	The search analogy for this characteristic is that user with growth mind are more exploratory learn from the existing experience. Accept challenges and obstacles to dig to the problem in details to find the best answer.

Derived from the above definition and analogies in Table 3 propose the following operational definitions for Growth mind-set:

- **Large number of query iteration.**
  Query iterations are alterations to the query by user. While the user has evolving information goals so there will be a lot of alteration to the query.
- **Use of large number of fixed jumps.**
  Fixed jumps are the types of query jumps used when there are no changes to the information goals user. If there is no changes to the user information goals and high number of jumps query are available the user has evoking information goals.
- **Use of large number of history.**
  History is an example of queries that are used before in sessions and between sessions. User with evolving information goals will use more repeats to refine the result.

3.2 Search Strategy

Search strategy is also divided into two types [16] namely, Cautious behavior and Risky behavior. Cautious users is described by Moulton Marston as analytical thinker, who has high standards, careful background research; focus on details, having realistic approach to solve the problem [1] and [18]. Marston work was ignored even it was the back bone of the prominent personality test tool called DiSC. The work is later on acknowledged by professor Irvine [4]. A cautious behavior is also considered self-disciplined, results-oriented, structured (organized) and slow mover [12].

3.2.1 Cautious behavior

To simplify cautious behavior Table 4 can provide a detailed explanation. From the right side of the table properties of cautious behavior can be seen and left side of the table clarify the analogy to information seeking.

Table 4: Cautious behaviors and its search analogy.

Cautious behavior	Analogy
Cautious behaviors are analytical thinker, have high standards, background research, focus on details, having realistic approach to solve the problem [18, 1, 4]. Cautious behavior make moves very carefully and Will organisable.	The search analogy for this characteristic is that cautious users will move around very carefully to choose a result. User will spend more time in single page before moving to a next page with new query. In selection of results user will carry out a very high query iteration and viewing large

	number of result pages to surf all opportunities.

The operational definitions are generated based on the search analogy to enable the application of the Cautious characteristics in user modeling are:

- **High number of query iteration.**
  The user will have more alteration to the queries.
- **More clicks in other pages compare to the first page.**
  The user will move around in the session and open more and more pages beyond first page.
- **Higher position link clicked in multiple pages.**
  The user will click on many links not only relying on the first links of the page.
- **View large number of result pages.**
  The user will view many pages correspond to the query to satisfy his information needs.
- **Spend long time per search iteration.**
  User will spend more time per session.

3.2.2 Risky behavior

In literature there is no perfect theory to explain risky behavior in contrary to cautious behavior. So it can be appeared opposite to cautious behavior. Right column of Table 5 explains analogy from the left column of the risky behavior.

Table 5: Risky behavior and its search analogy.

Risky behavior	Analogy
The Uses with Risky behavior will act oppositely to cautious behavior. So the selection of this kind of behavior is very superficial. The uses with risky behaviors provide no attention to details.	The search analogy for this characteristic is that Risky users will behave oppositely to Cautious users they run quickly around the pages and view low number of pages with low number iteration. Risky users doesn't concentrate on details there selection criteria is based on the surface knowledge and select results while browsing.

The operational definitions are derived from the conclusion of obtained analogy from risky behaviors.

- **Low query iteration.**
  The user will have less number of alterations to the queries.
- **Less number clicks in other pages compare to first page.**
  The user will rely on first page of the results.
- **Lower Position link clicked in multiple pages.**
  The user will click on the immediate link on the page for example on the first link of the page.
- **View small number of result pages.**
  The user will have small number of page viewing per session.
- **Spend short time per search iteration.**
  User will spend less time per session.

3.3 Evaluation threshold

The last set of the characteristics of the ISE model contains two types of result evaluation threshold, precise evaluation threshold and weak evaluation threshold. These characteristics involve decision making, so the study need to get some insight knowledge of decision making to strength the characteristics of the user model. According to Herbert A. Simon Nobel Laurent the process of decision making depends on the available information and understanding of the information in the required time [7]. Further decision makers are categories into two types' maximizers and satisfiers [24]. This section of the model will be will explained by these two types of decision makers in the below Table 6 and Table 7.

3.3.1 Precise Evaluation Threshold.

Precise evaluation threshold will be validated from the behaviors of maximizers as shown in Table 6. Left column of the table illustrates the behavior of maximizers and right column shows the obtained analogy from the behaviors.

Table 6: Maximizers and its search analogy to precise user.

Maximizers	Analogy
Maximizers take long time in decision making they are so specific having very high standards and check all available option to select and take decision when they are completely satisfied [7, 19, 24].	The search analogy for this characteristic is that Maximizers' care about source credibility which means they like high standards whether the information is reliable. This type of users will select result very carefully and with much iteration.

The operational definitions based on the search analogy enable the application of the precise characteristics in user modeling are:

- **High numbers of history.**
  The user will have high use of repetition queries in session and between sessions.
- **Larger numbers of clicks compare queries.**
  The user will have more clicks as compare to the amount of queries.
- **High query iteration.**
  The user will have more alteration to the queries.
- **Higher pages clicked.**
  The user will expend the search for long period and usage of more pages in session.
- **Search large number of iterations.**
  The user will perform more detail search e.g. more number of page viewing per iteration.

3.3.2 Weak Evaluation Threshold

Weak evaluation threshold will be validated from the behaviors of satisfiers as shown in Table 7. Left column of the table explain the behavior of satisfiers and right column shows the obtained analogy from the behaviors.

Table 7: Satisfiers and its search analogy.

Satisfiers	Analogy
Satisfiers take decision even the result have very little relevance and importance. They seem to think that an effort to find best result is waste of time. They would be happy if the result is good enough. [7, 19, 24].	The search analogy for this characteristic is that Weak users will select result with minimum relevance to the query. As long as there is some relevance the user is satisfied and will select results even with low resemblance.

The operational definitions based on the search analogy enable the application of the weak characteristics in user modeling are:

- **Low number of history.**
  The user will have less use of repetition query in session and between sessions.
- **Lower numbers of clicks compare to queries.**
  The user will have less number of page viewing and clicks as compare to the queries.
- **Low number of query iterations.**
  The user will have less number of alterations to the queries.
- **Lower number of pages clicked.**
  The user will open less number of pages during the search process.
- **Search small number of iterations.**
  The user will perform a very little during search process.
  View small number of pages during the search process.

4. EXPERIMENT SET UP

This section describes a preliminary experiment set up for validating the EISE model. Ten user search log data from Bing Search Engine is used, which contain a collection of 4231 queries average of 423 per user and 40217 results average of 4021 per user. The data log contain anonymous User id, time and date, query name, page number of search engine page, rank in one search engine page, URL, dwell time and click count. The Similar search log is used by a group of researcher to classify interaction features [11].

The extracted 24 key interaction features from the search log data justify the operational definition of each characteristics proposed in Section 3. The 24 features include, Average clicks on page number of search Result, Average Number of result pages viewed per query, Average Position of each result click on particular result page, Average Number of query per session, Average number of result clicks per session, Average time spend per session total number of query, total Empty result query, Average number of result clicks in single query, Total number of first link clicks, Average view time first link, Total number other link click, Average view time other links clicked, total number of repeat, Total Subset, total Super-set, Total number of Overlap, Total number of Back query, total number Back repeat queries between sessions, Total number of jump query , Total number Fixed jump query, Total number of new jump query between session, Total number of fixed new jump between session.

The whole process of data analysis was done manually as some of the key features require deeper analysis. From observing the data the below 10 types of query transitions are proposed. Five of them where already defined in the original ISE model for CBIR (Content based image retrieval) and the rest of other five were discovered during our data analysis.

• Repeat: Consecutive use of the same query [16].

• Subset: Subset of the previous query. [16].

• Super-set: The entire previous query with additional words [16].

• Overlap: Mix query with some words from previous query [16].

• Back: Same query used in a session but not consecutively.

• Back Repeat: Repeat of same queries in between sessions.

• Jump Query: New query within session and new information goal during the session.

• Fixed Jump Query: New query with fixed information goals during the session.

• New jump: New query with new information goals between sessions.

• Fixed New jump: New query with fixed information goals between sessions [16].

The above query iterations are divided into four groups, namely, history (repeat, back repeat, back); iteration (subset, super-set and overlap); Jumps (Jump and new jump); history plus iterations (fixed jump and fixed new jump).

5. EXPERIMENTAL RESULTS

This section report the data analysis results based on the methodology proposed in the above section. Applying the model on user search data log the result will be shown one by one characteristic wise. The above characteristics of the model overlap each other due to sharing of some operational definitions. The model divides the search process into three main categories Information goals, Search strategy and Evaluation threshold. The information seeking process looks simultaneous so there is a very thin line between these categories. To avoid confusion that from where each category starts if some of the operational definitions of each category shared then maximal results can be achieved from the data.

5.1 Results for information Goals

The selection criteria of users in fixed and evolving information is based on their performance in operational definitions. To further scrutinize the users for information goals the users are examine with low, medium and high performance in there operational definitions.

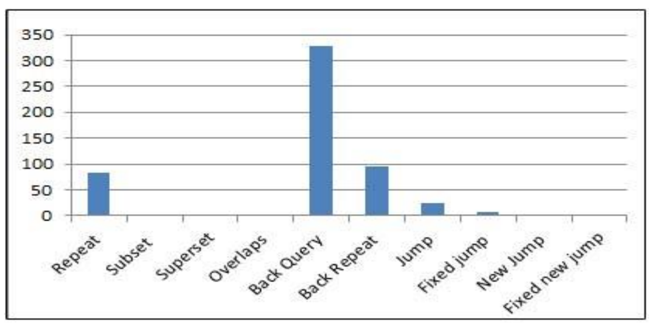

Figure 1: Example of number of with fixed information goal during search process.

5.1.1 *Fixed information Goals*

After analyzing the search data log the operational definitions of each characteristic explained in the model were judged beside the results. Figure 1 is an example of fixed information goals. According to operational definitions of fixed information goals user will have less number of query iterations (subset, super set and overlaps) an example of a user with no iterations in Figure 1 can be seen. The second operational definition is less fixed jumps (fixed jumps and fixed new jumps) which can be seen in Figure.1. There are a very low number of fixed jumps. The last operational definition is low history (repeat, back and back repeat) in this case the user has very high number of histories so this operational definition is not satisfied by Figure 1. Now this user qualifies two of the operational definition so the user is considered fixed in there information goals. In Table 8 shows all of the users that qualify each of the operational definitions related to fixed information goals but only User ID 7 and User ID 3 absorb maximum characteristics of fixed information goals.

Table 8: Fixed Users and there operational definitions.

Operational definitions	Users ID	Fixed Users ID
Less iteration	3, 7.	3, 7
Low fixed jumps	3, 4, 7, 9.	
Low history	3,	

5.1.2 *Evolving Information Goals*

Figure 2 is an example of a user with evolving information goals. According to the first operational definitions of evolving information goals, user will have high number of iterations (subset, super-set and overlaps). Fixed jumps (fixed jump and fixed new jump) are also consider iterations and history due to the same information goals. So Figure 2 shows a user with high number of iterations in combination with fixed jumps. The second operational definition is already explained and high fixed jumps can be seen in Figure 2. The last operational definition is High number of history (repeat, back and back repeat). While Figure 2 shows high number of repeats but low number of back and back repeats and fixed jumps also provides history. So not only relying on repeats than combine both repeats and fixed jumps user will have high number of histories. The overall performance of Figure 2 show that this is user had evolving information goals. Table 9 shows the overall performance of users with respect to operational definitions related to Evolving information goals. User ID 1, 2, 4, 5, 6, 8, 9 and 10 are identified to have evolving information goals.

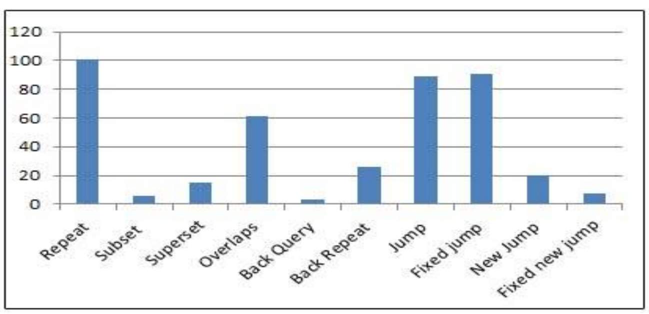

Figure 2: Example of number of with Growth information goal during search process.

Table 9: Evolving Users and operational definitions they belongs.

Operational definitions	Users ID	Evolving Users ID
High iteration	1, 2, 4, 5, 6, 8, 10.	1, 2, 4, 5, 6, 8, 9, 10.
High fixed jumps	1, 2, 5, 6, 8, 10.	
High history	1, 2, 4, 5, 6, 7, 8, 9, 10.	

5.2 Results for Search Strategy

Now search strategy is a combination of query iterations and other interaction behaviors. To visualize the interaction behaviors of users can be seen in Figure 1, Figure 2, Figure 3, Figure 4 and Figure 5, to understand the search strategy of the users

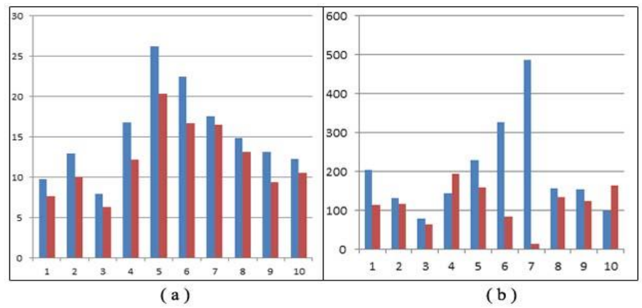

Figure 3: (a) Blue color is number of queries and Red color is result clicked (b) Blue color First link and Red color other link.

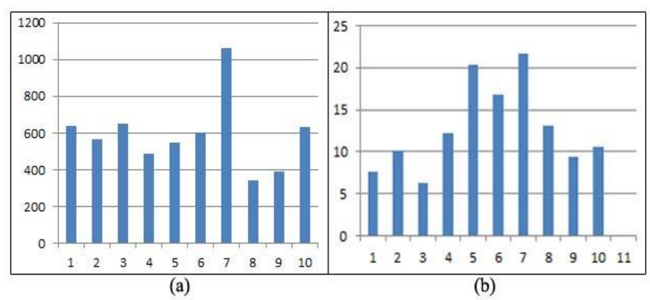

Figure 4: (a) Time per session (b) Clicks per session

5.2.1 *Cautious Strategy*

In Table 10 User ID 1, 4, 6 and 10 had cautious search strategy. These users are selected on the basis of maximum fulfillment of operational definitions to cautious users. First operational definition is high query iteration as explained in the above section with explanation of Figure 1 and Figure 2 that how the query iteration selection process work. In that context of User ID 1, 2, 4, 5, 6, 8 and 10 have high query iterations it means they have high number subset, super-set and overlaps. Figure 5 (a) illustrate the number of clicks per page this explains our second operational definition, high clicks in other pages compare to first page. So user ID 4, 6 and 10 belongs to this operational definition. Higher position link clicked is third operational definition of this characteristic can be seen in Figure 3 (b) average position of user clicks. Where User ID 2, 3, 4, 8, 9 and 10 had more clicks in other links rather than the first link of the page. The user in Figure 3 (b) cannot be compared with each other but they will be compared with their own clicks. Large number of pages viewed is fourth operational definition shown in Figure 5 (b) users with view number of pages where all users perform on average the same. Figure 4 (a) explain time spend by the user during search iteration and quantified by the average time spend by the users. 591 seconds per session is an average below this is low and above is high on this basis user qualifies the last operational definition. User ID 1, 3, 6, 7 and 10 spend more time than average so they are consider Cautious users. On the basis of these analyses the overall selection of users and selection of users to their corresponding operational definition can be seen in Table 10.

Table 10: Cautious users and there operational definitions

Operational definitions	Users ID	Cautious Users ID
High query iterations	1, 2, 4, 5, 6, 8, 10.	1, 4, 6, 10
High clicks in other pages compare to first page	4, 8, 10.	
Higher link Position clicked	2, 3, 4, 8, 9, 10.	
Large number of results pages viewed	0	
Spend long time	1, 3, 6, 7, 10.	

5.2.2 Risky Search Strategy

In Table 11 User ID 5 and 9 qualify the operational definitions for risky search strategy. First operational definition is less query iteration User ID 3, 7, 9 have less query iterations it means they have less number of subset, super-set and overlaps and an example of these iteration are shown in Figure 1 and Figure 2. Our second operational definition in Figure 5 (a) can be seen with fewer clicks in other pages compare to first page. So User ID 1, 2, 3, 5, 6, 7, 8 and 9 belongs to this operational definition. The third operational definition is low position link clicked can be seen in Figure 3 (b). Where user ID 1, 5, 6 and 7 had more clicks on first link so they had less effort. Less number of pages viewed is fourth operational definition shown in Figure.5 (b) explained before in the previous section that all of the users preform same on this operational definition. Figure 4 (a) explain time spend by the user during search iteration and 591 seconds per session is an average below this is low so User ID 2, 4, 5, 8 and 9. Spend less time than average so they are considering cautious with this operational definition.

According to the operational definitions of risky search strategy below selected users falls in this category shown in Table.11.

Table 11: Risky users and there operational definitions

Operational definitions	Users ID	Risky users
Low query iterations	3, 7, 9.	5, 9
less clicks in other pages compare to first page	1, 2, 3, 5, 6, 7, 8, 9,	
Lower links position clicked	1, 5, 6, 7.	
Small number of results pages viewed	0	
Spend short time	2, 4, 5, 8, 9.	

5.3 Results for Evaluation Threshold

The final phase of search process, at this stage user has to make a decision to select from one from the results.

5.3.1 Precise Evaluation Threshold

On the basis of data analysis User ID 4, 5 and 6 are precise in their selection of result to look in detail let see the data. The first operational definition can be explained with Figure 1 and Figure 2 as example and User ID 1, 2, 6, 7, and 8 are high history (Back, back repeat and also fixed jumps) users. Figure 3 (a) show the comparison of queries with clicks which is our second operational definition and according to the data User ID 4 and 5 belongs to this operational definition. The third definition can be explained with the example of Figure 1 and Figure 2 high query iterations(subset, super-set and overlaps) and User ID 1, 2, 4, 5, 6, 8, 9 and 10 falls in this operational definition. Figure 5 (a) illustrates users with higher pages clicks correspond to fourth operational definition and User ID 4, 8 and 10 belong to this operational definition. The users of last operational definition are selected on the basis of their average interaction (query and clicks) of a user in session. Average interactions is 12 clicks per session so above 12 clicks per session user belong to this definition with the User ID 4, 5, 6, 7 and 8 shown in Figure 4 (b). After analyzing the definition one by one the overall selection can be seen in Table 12.

Table 12: Precise users and there operational definitions.

Operational definitions	Users ID	Precise Users ID
High history	1, 2, 6, 7, 8.	4, 5, 6
Larger numbers of clicks compare queries	4, 5	
High query iterations	1, 2, 4, 5, 6, 8, 9, 10.	
Higher pages clicked	4, 8, 10	
Search large number of iterations	4, 5, 6, 7, 8.	

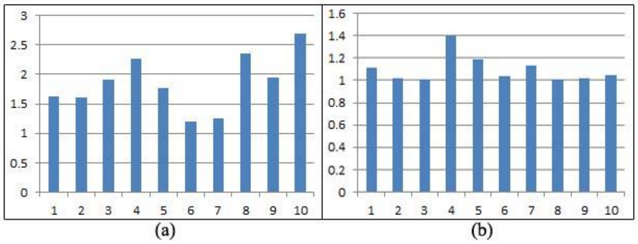

Figure 5: (a) Clicks per page (b) Pages viewed per query

5.3.2 Weak evaluation threshold

User ID 3, 9 and 10 have weak evaluation threshold according to their operational definitions. User ID 3, 4, 5, 9 and 10 are low history users and satisfy our first operational definition. According to Figure 3 (a) User ID 1, 2, 3, 6, 7, 8, 9 and 10, belongs to our second operational definition low clicks compare to queries. The third definition can be explained with the example of Figure 1 and Figure 2 low query iterations (subset, super-set and overlaps) and User ID 3 and 7 falls to this operational definition. In Figure 5 (a) User ID 1, 2, 3, 5, 6, 7 and 9 have low number of pages clicks and satisfy the forth operational definition. The average interaction (query and clicks) are 12 clicks per session. So the users perform below this interaction is low. User ID 1, 2, 3, 9 and 10 are low interaction users shown in Figure 3 (b). The users along with their operational definitions and selected users with weak evaluation threshold can be seen in table.13.

Table 13: Weak users and there operational definitions

Operational definitions	Users ID	Weak Precise
low history	3, 4, 5, 9, 10.	3, 9, 10
Lower numbers of clicks compare queries	1, 2, 3, 6, 7, 8, 9, 10,	
Small query iterations	3, 7.	
Lower pages clicked	1, 2, 3, 7,	
Search small number of iterations	1, 2, 3, 9, 10.	

6. DISCUSSION

The above results explain the validation of EISE model and show the existence of behaviors in user information seeking. Although the model is validated to an extent but still some concern can be raised about the operational definition. Up to now reduced numbers of operational definitions are used from the model because of available data limitation. Further study will extend the operational definitions to fully make use the psychology theories in the model. For example in search strategy an operational definition high standards/relevance occurs but from the current data we cannot judge this operational definition. In evaluation threshold operational definition high source credibility and high resemblance cannot be extracted from the data. To solve this problem new experiment in a control environment will take place to make use of all the operational definition and build a comprehensive model.

According to the operational definition users fall in the same categories but to fully absorb the characteristics of the model the user needs to qualify maximum operational definitions. For example in Search Strategy, Cautious User ID 1, 4, 6 and 10 and Risky User ID 5 and 9 can be seen but User ID 2, 3, 7 and 8 are missing they do not qualify maximum operational definition of the characteristics.. The same result can see in evaluation threshold that User ID 1, 2, 7 and 8 missing they do not qualifies maximum operational definitions. Now a detail investigation is needed to create a new characteristic for these users if they do not fall in the above categories

The results show the existence of characteristics independently from each other. Further study will also investigate the relation between these characteristics with each other's and the effect of information goals on search strategy and evaluation threshold and vice versa.

7. CONCLUSION

In this paper we present a user classification model for personalized information retrieval. In the previous studies models lack a theoretical background to personalize the search. In our study we build a model based on strong psychological theories to address human behavioral aspect and interpret those theories into information retrieval terminology. The proposed model is EISE (extended Information goal, Search strategy and Evaluation threshold) adapted from a model applied in CBIR (content based image retrieval). Enriching the model with the theories and applying on search data log we established the existing of those behaviors which we hypnotized in our model. On the basis of these behaviors we can build a personalized search to enhance user information seeking ability.

8. REFERENCES

[1] Azure. DISC based personality assessment. *http://www.amazureconsulting.com/files/1/74900324/DISCPastAndPresentAndWilliamMarston.pdf, 2011.* Accessed: 23 Dec. 2015.

[2] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short- and long-term behavior on search personalization. In *Proceedings of the 35th International Conference on Research and Development in Information Retrieval (SIGIR),* pages 185–194, 2012.

[3] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining, (SIGKDD),* pages 875–883, 2008.

[4] DiSC. DiSC Profile Explained. *https://www.discprofile.com/what-isdisc/overview/conscientiousness/, 2011*. Accessed: 23 Dec. 2015.

[5] C. S. Dweck. Mindset: The new psychology of success. Random House, New York, 2006.

[6] D. ELLIS. A behavioural approach to information retrieval system design. *Journal of Documentation,* 45(3):171–212, 1989.

[7] R. Gigerenzer, Gerd; Selten. Bounded Rationality: The Adaptive Toolbox. MIT Press, 2002.

[8] P. Ingwwrsen. Cognitive perspectives of information retrieval interaction: Elements of a cognitive ir theory. *Journal of Documentation,* 52(1):3–50, 1996.

[9] D. Jiang, K. W. T. Leung, J. Vosecky, and W. Ng. Personalized query suggestion with diversity awareness. In *Proceedings of the 37th International Conference on Research and Development (SIGIR)*, pages 400–411. 2014.

[10] D. Jiang, K. W.-T. Leung, L. Yang, and W. Ng. Query suggestion with diversification and personalization. *Knowledge-Based Systems*, 89:553–568, nov 2015.

[11] L. Jingfei, S. Dawei, P. Zhang, and Y. Hou. How different features contribute to the session search. In *Proceeding of 4th CCF Natural Language Processing and Chinese Computing (NLPCC),* pages 242–253, 2015.

[12] C. S. Jones and N. T. Hartley. Comparing correlations between four-quadrant and five-factor personality assessments. *American Journal of Business Education (AJBE),* 6(4), 2013.

[13] Y. Kim and W. B. Croft. Diversifying query suggestions based on query documents. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR),* pages 891–894, 2014.

[14] C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science,* 42(5):361–371, 1991.

[15] K. W. T. Leung and D. L. Lee. Deriving concept-based user profiles from search engine logs. *IEEE Transactions on Knowledge and Data Engineering,* 22(7):969–982, July 2010.

[16] L. Li, Z. Yang, B. Wang, and M. Kitsuregawa. Dynamic adaptation strategies for long-term and short-term user profile to personalize search. *In Proceedings of the Joint 9th Asia-Pacific Web and 8th International Conference on Web-age Information Management Conference on Advances in Data and Web Management,* pages 228–240, 2007.

[17] H. Liu, P. Mulholland, D. Song, V. Uren, and S. Ruger. Applying information foraging theory to understand user interaction with content-based image retrieval. In *Proceedings of the Third Symposium on Information Interaction in Context (IIiX), pages* 135–144, 2010.

[18] Z. Ma, G. Pant, and O. R. L. Sheng. Interest-based personalized search.*Transactions on Information Systems (ACM),* 25(1), feb 2007.

[19] W. M. Marston. Emotions of Normal People. Routledge, London, 1999.

[20] G. Nenkov, M. Morrin, A. Ward, B. Schwartz, and J. Hulland. A short form of the maximization scale: Factor structure, reliability and validity studies. *Judgment and Decision Making,* 3(5):371–388, 2002.

[21] G. Pasi. Issue in personalizing information retrieval. *Intelligent Informatics Bulletin*, 11(1):3–7, December 2010.

[22] P. Pirolli and S. Card. Information foraging in information access environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages (CHI),* 51–58, 1995.

[23] P. Pirolli and S. K. Card. Information foraging. *Psychological Review,* 106(6):643–675, 1999.

[24] G. H. Pyke. Optimal foraging theory: A critical review. *Annual Review of Ecology and Systematics*, 15:523–575, 1984.

[25] A. Schwartz, B. Ward, J. Monterosso, K. Lyubomirsky, S.and White, and D. R. Lehman. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology,* 83(5):1178–1197, 2002.

[26] D. Sontag, K. Collins-Thompson, P. N. Bennett, R. W. White, S. Dumais, and B. Billerbeck. Probabilistic models for personalizing web search. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)* pages 433–442., 2012.

[27] J. Teevan, S. T. Dumais, and E. Horvitz. Beyond the commons: Investigating the value of personalizing web search. In *Proceedings of the Workshop on New Technologies for Personalized Information Access*, pages 84–92, 2005.

[28] P. Thomas, A. Moffat, P. Bailey, and F. Scholer. Modeling decision points in user search behavior. In *Proceedings of the 5th Information Interaction in Context Symposium (IIiX),* pages 239–242., 2014.

[29] S. Verberne, M. Sappelli, K. Jarvelin, and W. Kraaij. User simulations for interactive search: Evaluating personalized query suggestion. In *Proceeding of the 37th Conference on IR Research (ECIR),* pages 768–690., April 2015.

[30] E. Werner and D. Hall. Optimal foraging and the size selection of prey by the bluegill sunfish (lepomis macrochirus). *Ecology*, 55(5):1042–1052, 1974.

[31] R. White, P. N. Bennett, and S. Dumais. Predicting short-term interests using activity-based search contexts. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM),* pages 1009-1018, Oct 2010.

[32] T. Wilson. On user studies and information needs. *Journal of Documentation,* 37(1):3–15, 1981.

[33] T. Wilson. Models in information behaviour research. *Journal of Documentation*, 55(3):249–270, 1999